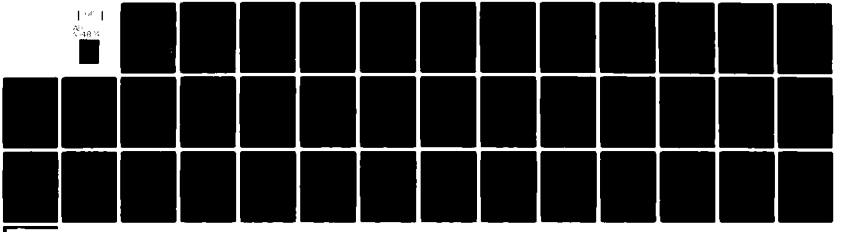
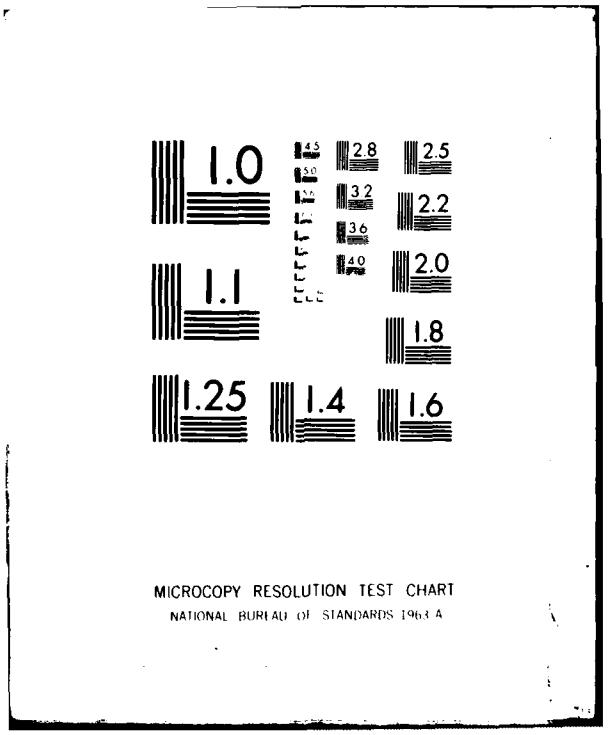
AD-A114 874 WASHINGTON UNIV SEATTLE LAB FOR CHEMOMETRICS F/G 7/4
QUANTITATIVE RESOLUTION OF FUSED CHROMATOGRAPHIC PEAKS IN GAS C--ETC(U)
MAY 82 M A SAHRAF, B R KOWALSKI N00014-75-C-0536
UNCLASSIFIED TR-23 NL



END
DATE FILMED
6 82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

(12)

OFFICE OF NAVAL RESEARCH

Contract N00014-75 C0536

Task No. NR 051-565

TECHNICAL REPORT NO. 23

QUANTITATIVE RESOLUTION OF FUSED CHROMATOGRAPHIC PEAKS IN GAS CHROMATOGRAPHY/MASS SPECTROMETRY

by

Muhammad Abdallah Sharaf and Bruce R. Kowalski

Prepared for Publication

in

Analytical Chemistry

University of Washington
Department of Chemistry
Seattle, Washington 98195

May 1982

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 23	2. GOVT ACCESSION NO.: AD-A11487
3. RECIPIENT'S CATALOG NUMBER: 4 --	
4. TITLE (and Subtitle): Quantiative Resoultion of Fused Chromatographic Peaks in Gass Chrmoatography/Mass Spectrometry	5. TYPE OF REPORT & PERIOD COVERED: Technical -- Interim
	6. PERFORMING ORG. REPORT NUMBER: --
7. AUTHOR(s): Muhammad Abdallah Sahraf and Bruce R. Kowalski	8. CONTRACT OR GRANT NUMBER(s): N00014-75 C0536
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Laboratory for Chemometrics, Department of Chem, University of Washington, Seattle, Washington 98195	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 051-565
11. CONTROLLING OFFICE NAME AND ADDRESS: Materials Sciences Division Office of Naval Research, Arlington, Virgina 22217	12. REPORT DATE: May 1982
	13. NUMBER OF PAGES: 35
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): --	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

to be published in Analytical Chemistry

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Curve Resolution
Generalized Inverse Method
Gas Chromatography-Mass Spectrometry (GCMS)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Curve Resolution and the Generalized Inverse Method are used to calculate the molar compositions of mixture mass spectra acquired during coelution of unseparated GC peaks. Quantiative resolution of the GC signal, independent of peak shape, results. The effects of noise, peak separation, a drifting has line and peak tailing are studied. No assumptions of peak shape or the presence of unique masses for the pure components are required.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Brief

Two procedures are developed to calculate the molar compositions of the composite spectra collected during coelution of partially separated components in GCMS.

Abstract

Curve Resolution and the Generalized Inverse Method are used to calculate the molar compositions of mixture mass spectra acquired during coelution of unseparated GC peaks. Quantitative resolution of the GC signal, independent of peak shape, results. The effects of noise, peak separation, a drifting base line and peak tailing are studied. No assumptions of peak shape or the presence of unique masses for the pure components are required.

INTRODUCTION

Combined gas chromatography-mass spectrometry (GCMS) has become an indispensible tool of analysis in many fields. Its popularity stems from the potential to analyze very complex samples in a matter of minutes and from the availability of simple to operate computer-controlled instruments (1). Normally GCMS data contain the information that is needed to positively identify and quantify major and minor constituents in samples(2,3). However serious problems arise when the eluting components are only partially separated and accurate qualitative and quantitative analyses become quite difficult. Fortunately, qualitative analysis is possible when the parent spectra of the chromatographically unresolved components are mathematically resolved(4).

Quantitative analysis of single component peaks by GC is a well documented and widely used technique offering general, selective and specific detection capabilities. It provides the analyst with an excellent linear dynamic range and reasonable precision and accuracy(5). With complex samples, however, various conditions may not allow complete resolution of all peaks which must be resolved before their areas or heights are measured for quantitative analysis.

Linear (geometric) and nonlinear (curve fitting) methods have been developed to resolve fused GC peaks(5,6). Linear methods suffer from being inaccurate and imprecise(6). Nonlinear methods are more accurate but they are difficult to employ as they require several parameters to be estimated by nonlinear parameter estimation techniques.

Poorly estimated initial parameters often lead to either divergence or convergence to the incorrect values(6). The accuracy of linear and nonlinear methods is determined by the shapes of the GC peaks, the mathematical peak shape function used and detector noise. The peak shape models adopted for a particular case may not apply if conditions, concentrations or columns are changed(7-11). Consequently, an algorithm or peak shape model chosen for one situation may have to be revised for another. Two techniques relying on curve fitting for resolving GCMS data have been described (12,13).

Smith et. al. and Blaisdell described two similar systems for quantitative and qualitative analysis of GCMS data(14,15). Both systems require libraries of previous analyses, a knowledge of masses unique to specific components and long computer time. These characteristics limit these methods to special purpose analyses. Knorr et. al. demonstrated an algorithm that models the time varying GC peak and extracts spectral information by a least square procedure without a dependence on unique masses(16). Successful application is dependent on the shape of the response surface whose minimum is being sought. Lundeen and Juvet related the GC response to a polynomial in concentration(17). The success of the method, however, requires that the analyst chooses the correct polynomial, knows the number of overlapping components in advance and guesses a good starting point to insure a quick convergence if any.

To overcome these problems the mass spectrometer has been used as a specific detector for the GC(18). The mass spectrometer is utilized in a selective ion monitor mode (SIM) or a limited ions monitor mode(LIM).

Both methods offer high sensitivity and selectivity. The techniques have been used successfully in quantifying chromatographically unresolved compounds(19-21). Complete structural information about the components of interest and other constituents in the sample is lost nevertheless. SIM and LIM are very explicit techniques; the analyst must obtain the spectra of the analytes prior to analysis and the absence of interferences must be assured. Additionally, the selected ions may not always be the most abundant ones and their response may not offer a useful linear dynamic range. While SIM and LIM have been successfully implemented in several situations, a fruitful analysis scheme must be "staged" in advance. Quantitation by SIM and LIM suffers from bad precision and larger variations (compared to GC) due to ion statistics(22-24). The variance of SIM quantitation can be twice that of employing a flame ionization detector(25). The attractiveness of SIM and LIM is attributed to speed of analysis and high selectivity. When unique masses exist and they are known, there is no need to resolve GC peaks before quantitation. A system for GCMS data reduction relying on SIM and LIM detection has been described(26).

Since GC offers the linearity, accuracy and precision needed for complex sample analysis, it is desirable to extend its capability to quantify overlapping peaks without relying on limiting assumptions (i.e. specific ions). This paper introduces one method and illustrates and compares two methods, each depending on multivariate curve resolution(4,27), for resolving GC peaks. Both methods are fast, accurate and independent of the shape of the GC peaks. Most importantly, they do not require

peak shape modeling or the estimation of any peak shape parameters. For simplicity, the methods will be illustrated as applied to chromatographically unresolved binary mixtures. A generalization to more complex multicomponent systems is in progress.

The two methods offer the analyst a direct solution to the peak deconvolution problem and ensure that the complete instrumental resolution is realized. Moreover, quantitation accuracies can actually be better than instrument response precision due to a degree of signal averaging inherent to multivariate statistical analysis.

EXPERIMENTAL

Instruments

Experimental data were obtained using a Hewlett-Packard 5985 GCMS. The system consists of an HP 5840 microprocessor controlled GC, a jet separator and a hyberbolic quadrupole mass spectrometer.

Material

Analytical grade reagents with a boiling range of 0.5-1.0°C were used without further purification.

Procedure

The mixtures were injected in a 6 ft packed glass column. The packing material used was 3% SE-30 on 100/120 mesh Gas Chrom Q support. The resolution of the effluents was varied by adjusting the flow rate and the temperature of the column accordingly. The mass spectra were scanned at a rate of 1 scan per second covering the range from 15-150 M/z.

Computer Programs

A Fortran program that can correctly determine the number of components under a GC peak and then performs qualitative and quantitative resolution is available from Infometrix, Inc., P.O. Box 25808, Seattle, WA 98125. The analysis of a data matrix containing 300 entries (e.g. 10x30) requires 40K byte on a Z-80 micro-processor based micro-computer compiled with micro-soft Fortran.

THEORY

The Generalized Inverse Method. When two components coelute from the GC the collected mass spectra are assumed to be linear combinations of the components pure mass spectra. At every point in time, during the emergence of the combined peaks, the eluting mixture is composed of $\alpha(t)$ mole fraction of the first component and $\beta(t)$ mole fraction of the second component. The intensity of the GC responses are also considered to be non-negative linear combinations of the responses of the pure components. The combinations in both cases are the same. If N mass spectra are collected by scanning I M/z signals, the experimental mass spectra constitute a data matrix, $\underset{\sim}{X}$, and are represented by the following matrix equation:

$$\underset{\sim}{X}_{N,I} = \underset{\sim}{C}_{N,2}\,\underset{\sim}{P}_{2,I} \qquad (1)$$

where c_{kj} is the mole fraction of the j^{th} component in the k^{th} scanned spectrum (the k^{th} eluting mixture), and p_{jl} is the intensity of the l^{th} signal in the j^{th} pure component. If the pure spectra are either known or can be estimated, an assumption that can be verified by statistics, the solution band widths(4), or library search, then

$$\underset{\sim}{C} = \underset{\sim}{X}\underset{\sim}{P}^{T}(\underset{\sim}{P}\underset{\sim}{P}^{T})^{-1} \qquad (2)$$

where $\underset{\sim}{P}^{T}$ is the transpose of the matrix $\underset{\sim}{P}$ and $\underset{\sim}{P}^{T}(\underset{\sim}{P}\underset{\sim}{P}^{T})^{-1}$ is the so called generalized inverse of $\underset{\sim}{P}$. This solution for $\underset{\sim}{C}$ is nothing new. It is simply the multivariate least squares solution. It is used here as a standard benchmark for the method described in the next section.

The time varying response, i.e. TIC, FID, at every point in time, i, when a mass spectrum is scanned is resolved by partitioning the total signal according to c_{i1} and c_{i2}. If the ratio of c_{i1} to c_{i2} is denoted by R_i, then the contributions of the first component, $S_{1,i}$, and the second component, $S_{2,i}$, to the total GC signal, $S_{tot,i}$, are

$$S_{1,i} = [R_i/(1 + R_i)]S_{tot,i} \quad (3)$$

$$S_{2,i} = [1/(1 + R_i)]S_{tot,i} \quad (4)$$

If R_i can be estimated for each mixture spectrum, then quantitative resolution in the time domain is straightforward.

The Curve Resolution Method. Factor analysis of the data matrix, $\underline{X}$, yields two principle eigenvectors, $\overline{V}_1$ and $\overline{V}_2$, containing all the variance associated with the two chemical components(4). Each experimentally scanned mass spectrum, $\overline{MS}_i$, is represented by a point, (a_i,b_i), in a two dimensional (factors' loadings) space. Coordinates (a_i,b_i) for each point are given by

$$a_i = \overline{MS}_i.\overline{V}_1 \quad (5)$$

$$b_i = \overline{MS}_i.\overline{V}_2 \quad (6)$$

After normalization to unit area, the points representing all N measured spectra lie on a straight line in the factors' loadings space. This is shown in Figure 1 with N = 4.

The shaded areas in Figure 1 represent the two solution bands that contain the parent spectra. In the eigenvector domain, each solution band is defined by an "outer" and an "inner" spectrum. Thus, points (a_{u1},b_{u1}) and (a_{u2},b_{u2}) represent the outer spectra of

the first and second solution bands respectively. Similarily, (a_1,b_1) and (a_4,b_4) represent the inner spectra of the solution bands. In the spectral domain, however, the outer and inner spectra will, for convenience, be referred to as the upper and lower edges of the solution bands respectively. The two points designated by (a_{u1},b_{u1}) and (a_{u2},b_{u2}) in Figure 1 correspond to the spectra of the upper edges of the solution bands(4). The points corresponding to the spectra of the lower edges of the solution bands in Figure 1, (a_1,b_1) and (a_4,b_4), represent the "purest" mass spectra scanned during elution. Note that these points do not always correspond to the first and last scanned mass spectra. Other points in between represent non-negative linear combinations of these spectra. Let m and n symbolize the purest two spectra and let i be any middle point representing a mixture of α' mole fraction of m and β' mole fraction of n. The following relationships hold:

$$\overline{MS}_i = \alpha' \overline{MS}_m + \beta' \overline{MS}_n \tag{7}$$

$$\overline{MS}_m = a_m\overline{V}_1 + b_m\overline{V}_2 \tag{8}$$

$$\overline{MS}_n = a_n\overline{V}_1 + b_n\overline{V}_2 \tag{9}$$

$$\overline{MS}_i = a_i\overline{V}_1 + b_i\overline{V}_2 \tag{10}$$

Substituting eqs. 8, 9 and 10 for $\overline{MS}_m$, $\overline{MS}_n$ and $\overline{MS}_i$, respectively, in eq. 7 gives

$$a_i\overline{V}_1 + b_i\overline{V}_2 = (\alpha'a_m + \beta'a_n)\overline{V}_1 + (\alpha'b_m + \beta'b_n)\overline{V}_2 \tag{11}$$

The above vector equation implies that

$$a_i = \alpha' a_m + \beta' a_n \tag{12}$$

and

$$b_i = \alpha' b_m + \beta' b_n \tag{13}$$

The Euclidean distance between point m and point i, d_{mi}, in the factors' loadings space is given by

$$d^2_{mi} = (a_m - a_i)^2 + (b_m - b_i)^2 \tag{14}$$

Substituting for a_i and b_i from eqs. 12 and 13, eq. 14 can be rewritten as

$$d^2_{mi} = \beta'^2[(a_m - a_n)^2 + (b_m - b_n)^2] \tag{15}$$

Similarly,

$$d^2_{ni} = \alpha'^2[(a_m - a_n)^2 + (b_m - b_n)^2] \tag{16}$$

Therefore,

$$d_{mi}/d_{ni} = \beta'/\alpha' \tag{17}$$

This proves that the ratio of the distance between points m and i to the distance berween points n and i is the same as the ratio of the mole fraction of n in i to the mole fraction of m in i. If m and n represent the spectra of the pure components, an assumption that can be verified as mentioned above, then α' and β' are also the true mole fractions α and β.

If m and n do not represent the spectra of the pure components, then α' and β' are only estimates of the true mole fraction α and β. The adequacy of these estimates is determined by the chromatographic

resolution as well as the shapes of the peaks. The better the resolution the closer α' and β' to α and β respectively.

If we let

$$R_i = d_{mi}/d_{ni} \tag{18}$$

Then the total chromatographic signal at point i, $S_{tot,i}$, is resolved into the contribution of the first component, $S_{m,i}$, and that of the second component, $S_{n,i}$, according to

$$S_{m,i} = [1/(1 + R_i)]S_{tot,i} \tag{19}$$

and

$$S_{n,i} = [R_i/(1 + R_i)]S_{tot,i} \tag{20}$$

RESULTS AND DISCUSSION

Simulated Data. The coelution of α-pinene and pinane, two structurally similar compounds, is represented by a gamma and a Gaussian peak shape profile respectively. The choice was made to examine the effects of tailing. The symmetric Gaussian profile is superimposed on the "tail" of the unsymmetric gamma function. Various chromatographic resolution situations are simulated by moving the Gaussian profile along the time axis. The separation of the two peaks can be expressed as the distance between the two maxima in units of σ, the standard deviation of the Gaussian peak, which is also equal to the variance of the gamma peak profile (σ=0.5). Mass spectra sampled as a function of time are assumed to be linear combinations of the literature mass spectra of pure components(28). The linear combinations are determined by the relative heights of the theoretical profiles.

Figures 2 and 3 show the results of resolving the convoluted chromatographic peaks at separations of 0 and 4 σ respectively. The calculated curves are identical to the theoretical ones. Results of the Curve Resolution Method (CR) and the Generalized Inverse Method (GI) with the first and last scanned spectra used as standards are the same. The situation shown in Figure 2 represents a chromatographic resolution of zero. The procedures, however, rely on spectral resolution and chromatographic resolution. In this case, spectral resolution is achieved because the peaks have different shapes.

When noise is added to the simulated mixture mass spectra (a more realistic situation), the applications of CR and GI yield different

results. Random noise of 1%, 3%, 5%, 7% and 10% was added. The GC peaks are then resolved by CR and GI and the areas under the calculated curves are compared to the true areas. When noise is present, the first and last acquired spectra are badly affected because of the low total intensity. An advantage of CR over GI can be quickly realized as CR is able to indicate the least contaminated spectra. If GI is to be employed, the analyst has to examine, by library searches or inspection, several spectra in order to determine the purest acquired ones. In contrast CR eliminates the need for such validation steps necessary for data reduction. In the following analysis, the least contaminated spectra indicated by CR are used as standards in both CR and GI. Tables I and II show the average errors of 30 random perturbations obtained by both methods at separations of 0 and 4 σ respectively. CR is more "accommodating" to noise than GI. Errors obtained by application of GI can be twice as high as those obtained by CR. This is due to the fact that CR calculates the GC curves in an eigenvector space. Factor analysis is a signal averaging procedure. It is a powerful statistical means of minimizing the effects of random variations(29). In contrast, the curves calculated by GI are subject to an error propagation that is dependent on the condition number of the data matrix(30). As chromatographic separation increases, the effects of random noise are minimal.

When the noise level is very high (>10%) CR is disadvantaged because the noise becomes a "factor". The averaging of systematic changes is inadequate in accounting for the total variance and significant errors are introduced. At optimal instrumental operating conditions where random variations are less than 10% CR becomes the method of choice.

The composition of the eluting mixture was varied by changing the relative areas under each peak. Two typical cases are investigated where the two maxima are separated by 4 σ. In the first case, the Gaussian peak is preceded by a smaller gamma peak. The noise-free

data are, again, perturbed by a random noise of 1% to simulate a real-life laboratory situation. Table III summerizes the errors in calculating the areas under both curves by CR. The reported deviations are the averages of 30 different perturbations. The Gaussian profile does not cause problems as it predominates the total chromatogram. The combined signal is accurately resolved. The second case is the opposite of the first one. The gamma distribution is followed by a smaller Gaussian shaped peak. Table IV lists the average errors in calculating the areas under the GC peaks by CR. The errors shown are the averages of 30 1% randomly perturbed data. As the tailing peak predominates, the Gaussian profile becomes completely "buried" underneath it. No uncontaminated mass spectral scans are obtained for the component represented by the Gaussian profile because of low S/N of the peak. This "tailing" prevents accurate reproduction of the GC signal at lower relative compositions. At comparable ratios the GC signals are resolved successfully.

A drifting base line during the elution of a GC single peak can prevent accurate quantitative analysis. The effects of a drifting base line are not restricted to detector noise. A bleeding column, for example, introduces extra mass spectral signals that preclude correct identification of the eluting component and prevents accurate quantitation in the chromatographic domain. The information pertaining to a shift in the base line and/or a column bleed is contained in the intensity and spectral data. By factor analyzing these data, CR can successfully separate these effects from the information associated with the eluting peak. Drifting base lines can be treated as interfering components using the curve resolution approach and can be resolved from the GC signal.

Two kinds of drifting base lines are simulated here. The first

is a linear base line of the form $f(t) = k_1 t$ and the second is a nonlinear base line of the form $f(t) = k_1 t + k_2 t^2$. The value of t varied sequentially from 1 to N, where N is the total number of mass spectra acquired. In both cases the base line drift accounted for 3% of the total peak area. The simulated data are perturbed by 1% and 3% random noise. Table V shows the results of resolving the single peak and the base line by CR. The deviations in Table V are the averages of 30 random perturbation. The peak and the base line in both cases are successfully resolved.

Experimental Data. Figure 4 shows the coelution of n-heptane and methylcyclohexane at three different degrees of resolution. The chromatographic response, Total Ion Current (TIC), was resolved into two peaks, one for n-heptane and the other is for methylcyclohexane, by CR and GI. Methylcyclohexane and n-heptane have specific signals at M/z=98 and M/z=100 respectively. These specific signals were regarded as "internal standards" because their responses are completely resolved regardless of the chromatographic resolution. In order to make the problem much more difficult for CR, these specific ion signals were exculded from the data matrix when the GC peaks were resolved. The areas of the resolved GC peaks were then compared to the areas of the specific mass chromatograms (MCs) in all three cases shown in Figure 4. Table VI lists the areas of the resolved GC peaks and those of the MCS of the specific, but not the most abundant, mass spectral signals for n-heptane and methylcyclohexane. The analytical reliability of the areas of the resolved GC peaks is evaluated by correlating them to the areas of the specific MCs. The Fisher transformation is used to establish the 99% confidence intervals of the correlation coefficients (31).

Tables VII and VIII list the regression parameters and the corelation coefficients and their 99% confidence intervals for the areas of the specific MCs and the resolved GC peaks for n-heptane and methylcyclohexane respectively.

The high correlations between the areas of the resolved GC peaks and the areas of the always resolved MCs are evident. There is no loss of information upon application of CR and GI to the convoluted GC signal. The regression parameters for each component remain virtually unchanged even though the chromatograhic resolution is worse by a factor of 5 (case c vs. case a). This and the high correlations indicate that in all three cases the GC peaks are accurately resolved and that the resolved GC signals are as analytically reliable as the specific MCs. The potential of multivariate curve resolution is thus evident as it allows the GC response to furnish information equivalent, but with better precision, to that of a specific detector. Note that the specific signals were excluded from the data matrix before analysis. This is a clear advantage of the curve resolution method over SIM, LIM and other methods that rely on specific MCs.

Additionally, CR requires the analyst to collect multivariate data. This improves the analytical signal, reduces the effects of random background variations and detector noise and increases the amount of information acquired during the experiment. Moreover, CR (as shown above) does not require specific signals to be present. A very important feature of CR is that it indicates the purest scanned mass spectra. Since these spectra are not always the first and last scanned spectra, CR presents a valuable tool for choosing the mass spectra most suitable for library searches.

In the present example, both CR and GI have performed equally well. The data do not show any significant differences in the results. A t-test shows that the differences in the areas of the n-heptane peaks

as calculated by CR and GI are insignificant at the 0.05 significance level. The same is true for the methylcyclohexane peaks. However, CR is generally preferred because of its ability to accommodate experimental noise. In GI the accuracy is determined to a large extent by the condition number of the data matrix; the more similiar the parent spectra, the higher the condition number.

The accuracy of curve resolution depends upon at least three limitations. First, the inherent instrumental resolving power imposes a limit that is not possible to exceed via data analysis methods. A good example is the extreme case of two components eluting from the column at exactly the same time with the same peak profiles: All mixture spectra are identical in this case and resolution is impossible.

The next limiting factor is the uniqueness of the spectra of the eluting compounds. If the spectra contain at least one unique mass each (it is important to realize that the unique masses are not known a priori), then the pure spectra will lie on the upper edges of the solution bands(4). If the spectra of the eluting compounds are quite similiar and contain no unique masses, the upper edges of the solution bands will <u>not</u> represent the true parent spectra. This deviation will be reflected in the band widths of the solution bands. In this case finite band widths will be obtained regardless of the chromatographic resolution (see case III below).

The angle related to spectral uniqueness is formed by the intersection of the upper edges of the solution bands at the orgin in Figure 1. When the spectra of the pure components are identical

this angle is zero and no resolution is possible. At the other extreme, the angle approaches 90° as the spectra are more dissimilar. If all signals in the parent spectra are specific, this angle becomes 90°.

The third limitation is the rate of spectral data acquisition. It is desirable to collect as many mass spectra as possible during the time the compounds are eluting. The three benefits that occur are improved signal to noise ratio (more spectra are "averaged"), better integration accuracy (more points are available on the peak profile) and the acquisition of mixture spectra (the two spectra that represent the lower edges of the solution bands) that are closest to pure spectra. In practice the speed of scanning is determined by the instrumental background and the desired quality of the acquired mass spectra. The following section describes the best procedures for qualitative and quantiative analyses for possible situations that the analyst, or the computer program, may face.

Case I. The widths of the solution bands are equal to zero: This is the most favorable situation. It is an indication that specific signals are present and that pure mass spectra have been acquired. The analyst can use the pure spectra for qualitative analysis and the factor loadings, i.e. (a_1,b_1) and (a_4,b_4) in Figure 1, to calculate the molar compositions. No assumptions are required.

Case II. The widths of the solution bands are not equal to zero and specific signals are known: This is an indication that all acquired spectra are mixture mass spectra and none of them should be used as pure spectra for library searches and other subsequent processing.

The upper edges of the solution bands are the parent mass spectra and should be used for qualitative analysis. For quantitative analysis the points defining the upper edges of the solution bands, i.e. (a_{u1},b_{u1}) and (a_{u2},b_{u2}) in Figure 1, should be used to calculate molar compositions.

Case III. The widths of the solution bands are not equal to zero and specific signals are not known: This is the most challenging situation. When specific signals are not present finite solution band widths will result even if the chromatographic resolution is high. The analyst can use the average spectra for library searches after excluding any signal with an unacceptably large range(4). It is recommended, however, that the lower edges of the solution bands (the purest acquired spectra) be first examined to decide whether or not they are indeed the parent spectra or at least good estimates. As mentioned earlier, this can be accomplished by examining the chromatogram or the band widths themselves. The evaluation of the lower edges of the solution bands can be very advantageous. If they are found to adequately estimate the parent spectra then they can be used for qualitative analysis. Quantitative analysis is accomplished by using their loadings (i.e. (a_1,b_1) and (a_4,b_4) in Figure 1) to calculate molar compositions. If the lower edges of the solution bands are not found to be adequate estimates of the parent spectra, the analyst can calculate intervals for the molar compositions using both the upper and the lower edges of the solution bands. The GC peak is, consequently, resolved twice and the chromatographic response of each component is represented by a region that contains the true

chromatogram and establishes an interval for its area.

In the above treatment of experimental data the specific signals were removed from all of the mass spectra. As a result, finite band widths were obtained when CR was applied to all the three cases shown in Figure 4. The solution bands are used to calculate an interval for the area of each individual chromatogram. Table IX lists intervals of areas of the benzene and the methylcyclohexane signals for each case shown in Figure 4. The mean of each interval shown in Table IX is a good estimate (within less than 3%) of the corresponding value listed in Table VI. The upper and lower bounds of the areas of the GC signals define the concentration ranges of the individual components in the sample. Calculating intervals for the individual GC peaks is, of course, a "last resort" solution. It is important to realize that the analyst can use CR to estimate intervals for the concentrations of the components even when they are completely inseparable (the method of linear least squares regression can not be used) and no information about their mass spectra is available (LIM, SIM and curve fitting of MCs are inapplicable).

CONCLUSION

When the analyst uses the multivariate curve resolution method described in this paper, the full amount of useful information that the instrument can deliver is realized.

Accurate qualitative and quantitative analyses by GCMS are possible in the absence of complete chromatographic resolution. With a separation equivalent to only one mass spectral scan cycle (only one pure mass spectrum is acquired for each component), CR offers a powerful resolution tool that does not require limiting assumptions. Presently, the method is applicable to binary mixtures. Work in progress involves application of the methods to other chromatographic-spectrometric systems (e.g. LC-UV) and a generalization to more complex cases.

Acknowledgement:

The stimulating dicussions and critiques of Maynarhs da Koven are sincerely acknowledged.

Credit:

This research was supported in part by the Office of Naval Research. M.A.S. gratefully acknowledges the generous support of the University of Petroleum and Minerals, Dhahran- Saudi Arabia.

LITERATURE CITED

(1) Brooks, J.W. and Middleditch, B.S. Mass Spectrom, 1979, 5, 142.

(2) Burlingame, A.L.; Baillie, T.A.; Derrick, P.J. and Chizhov, O.S. Anal. Chem. 1980, 52, 214R.

(3) Cram, S.P.; Risby, T.H.; Field, L.R. and Yu, Wei-Lu Anal. Chem., 1980, 52, 324R.

(4) Sharaf, Muhammad A. and Kowalski, Bruce R. Anal. Chem. 1981, 53 518.

(5) Novak, J. "Quantitative Analysis by Gas Chromatography", Dekker, New York, 1975.

(6) Caesar, F. "Topics in Current Chemistry", Spinger-Verlag, 1973, 39, 139.

(7) Hancock, H.A.; Dahm, L.A. and Muldoon, J.F. J.Chromatogr.Sci., 1970, 8, 57.

(8) Anderson, A.H.; Gibb, T.C. and Littlewood, A.B. J.Chromatogr. Sci., 1970, 8, 640.

(9) Baudisch, J.; Papendick, H.D. and Schloder, V. Chromatrographia, 1970, 3, 469.

(10) Novak, J.; Petoovic, K. and Wicar, S. J.Chromatogr., 1971, 55 221.

(11) Kishimoto, K.; Miyauchi, H. and Mush, S. J. Chromatogr. Sci., 1972, 10, 27.

(12) Dromey, R.G.; Stefik, M.J.; Rindfleisch, T.C. and Duffield, A.M. Anal. Chem., 1976, 48, 1368.

(13) Blaisdell, B.E. and Sweeley, C.C. Anal. Chim. Acta, 1980, 117, 1.

(14) Smith, D.H.; Achenbach, M.; Yeager, W.J.; Fitch, W.L. and Rindfleisch, T.C. Anal. Chem., 1977, 49, 1623.

(15) Blaisdell, B.E. Anal. Chem, 1977, 49, 180.

(16) Knorr, F.J.; Thorsheim, H.R. and Harris, J.M. Anal.Chem., 1981, 53, 821.

(17) Lundeen, J.T. and Juvet, R.S. Anal. Chem, 1981, 53, 1369.

(18) Fenselau, C. Anal. Chem., 1977, 49, 563A

(19) Sweeley, C.C.; Elliot, W.H.; Fries, I. and Ryhage, R. Anal. Chem, 1966, 38, 1549.

(20) Eichelberger, J.W.; Harris, L.E. and Budde, W.L. Anal. Chem, 1974, 46, 227.

(21) Murphy, R.C.; Hattox, S.E. and Helbig, H.R. Biomed. Mass. Spectrom., 1978, 5, 444.

(22) Sweeley, C.C.; Young, N.D.; Holland, J.F. and Gates, S.C. J. Chromatogr. 1974, 99, 507.

(23) Falkner, F.C.; Sweetman, B.J. and Watson, J.T. Appl. Spectrosc. Rev., 1975, 10, 51.

(24) Budde, W.L. and Eichelberger, J.W. J. Chromatogr., 1977, 134, 147.

(25) Sheehan, M. and Hertel, R.A. Clin.Chem., 1976, 22, 1167.

(26) Gates, S.C.; Misko, M.J.; Ashendel, G.L.; Young, N.D.; Holland, J.F. and Sweeley, C.C. Anal. Chem., 1978, 50, 433.

(27) Lawton, W.H.; Sylvester, E.A. Technometrics 1971, 13, 617.

(28) Stenhagen, E.; Abrahamson, S.; McLafferty, F.W., Eds. "Atlas of Mass Spectral Data"; Wiley:New York, 1969.

(29) Malinowski, E.R. Anal. Chem., 1977, 49, 606.

(30) Jochum, C.; Jochum, P. and Kowalski, Bruce R. Anal. Chem, 1981, 53, 85.

(31) Fisher, R.A. Metron, 1921,1, 3.

Table I. Percent error in the calculated areas of α-pinene and pinane chromatograms at 0σ separation.

Profile	Noise									
	1%		3%		5%		7%		10%	
	CR	GI	CR	GI	CR	GI	CR	GI	CR	GI
Gamma	0.51	0.99	1.30	2.83	1.50	2.17	3.37	4.51	3.73	4.94
Gaussian	0.50	1.01	1.30	2.81	1.43	2.30	3.39	4.53	3.61	5.08

Table II. Percent error in the calculated areas of α-pinene and pinane chromatograms at 4σ separation.

Profile	Noise									
	1%		3%		5%		7%		10%	
	CR	GI	CR	GI	CR	GI	CR	GI	CR	GI
Gamma	0.29	0.40	0.68	1.00	1.60	2.42	1.42	2.12	2.5C	2.84
Gauss	0.30	0.41	0.68	0.98	1.63	2.39	1.42	2.08	2.45	2.84

Table III. Percent error in the calculated areas of a Gaussian profile (A_2) preceeded by a small gamma profile (A_1) at 1% Noise and 4σ separation

Profile	$A_1:A_2$ 1:1	1:4	1:20	1:100
Gamma	0.29	0.78	0.41	0.35
Gauss	0.30	0.09	0.02	0.00

Table IV. Percent error in the calculated areas of a gamma profile (A_1) followed by a small Gaussian profile (A_2) at 1% noise and 4σ separation.

$A_2:A_1$

Profile	1: 1	1: 4	1:20	1:100
Gamma	0.29	0.33	0.29	0.43
Gauss	0.30	1.33	5.66	42.81

Table V. Percent error in the calculated areas of a single gaussian peak and a drifting base line at 1% and 3% noise.

Noise	peak	linear base line	nonlinear base line
1%	0.03	0.26	0.27
3%	0.05	0.20	0.21

Table VI. Areas (arbitrary units) of resolved TIC peaks and specific Mass Chromatograms of n-heptane (normal heptane) and methylcyclohexane.

case	MC		GI		CR	
	M/z=100	M/z=98	normal heptane	methylcyclo-hexane	normal heptane	methylcyclo-hexane
a	29243	54603	195170	121741	195185	121756
b	18686	38777	125210	86388	125155	86418
c	11288	26341	75858	59152	75640	59068

Table VII. Correlation between MC and CR and between MC and GI for the areas of the n-heptane signals.

	MC and CR	MC and GI
Regression Coefficient, b (standard deviation)	0.150^{a} (1.00×10^{-3})	0.149^{b} (7.07×10^{-4})
Correlation Coefficient r	1.0000	1.0000
0.99 Confidence interval for ρ	[0.9998 - 1.0000]	[0.9996 - 1.0000]

[a]model: $A_{MC} = \beta A_{CR}$

[b]model: $A_{MC} = \beta A_{GI}$

Table VIII. Correlation between MC and CR and between MC and GI for the areas of the methylcyclohexane signals.

	MC and CR	MC and GI
Regression Coefficient, b (standard deviation)	0.448^{a} (1.58×10^{-3})	$0.448^{b}(2.35 \times 10^{-3})$
Correlation Coefficient, r	1.0000	1.0000
0.99 Confidence interval for ρ	[0.9994 - 1.0000]	[0.9983 - 1.0000]

[a] model: $A_{MC} = \beta A_{CR}$

[b] model: $A_{MC} = \beta A_{GI}$

Table IX. Intervals for the areas (arbitrary units) of the resolved TIC peaks of n-heptane and methylcyclohexane.

Case	n-heptane	methylcyclohexane
a	[190442 - 202620]	[114174 - 126500]
b	[120811 - 132005]	[79568 - 90762]
c	[70524 - 83947]	[50776 - 64199]

Figure Captions

Fig. 1: Four pairs of loadings representing four mass spectra in the factor loadings space. The points (a_1,b_1) and (a_4,b_4) define the lower edges of the solution bands. Similarly the points (a_{u1},b_{u1}) and (a_{u2},b_{u2}) designate the upper edges of the solution bands.

Fig. 2: Simulated coelution of α-pinene and pinane at 0.0σ separation. The solid lines are the theoretical profiles and their sum. The ○s amd the ■s are the calculated individual responses.

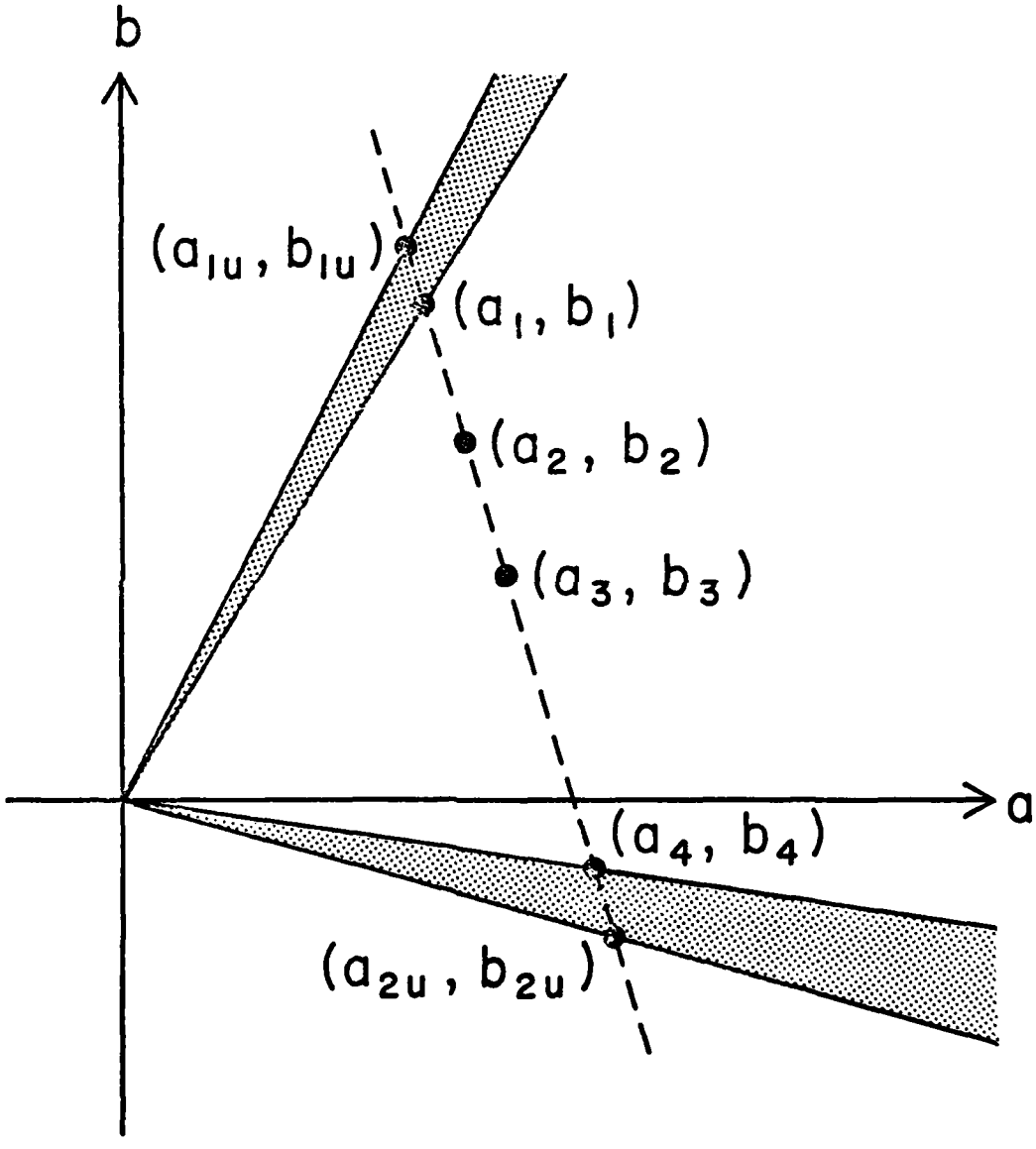
Fig. 3: Simulated coelution of α-pinene and pinane at 4.0σ separation. The solid lines are the theoretical profiles and their sum. The ○s and the■s are the calculated individual responses.

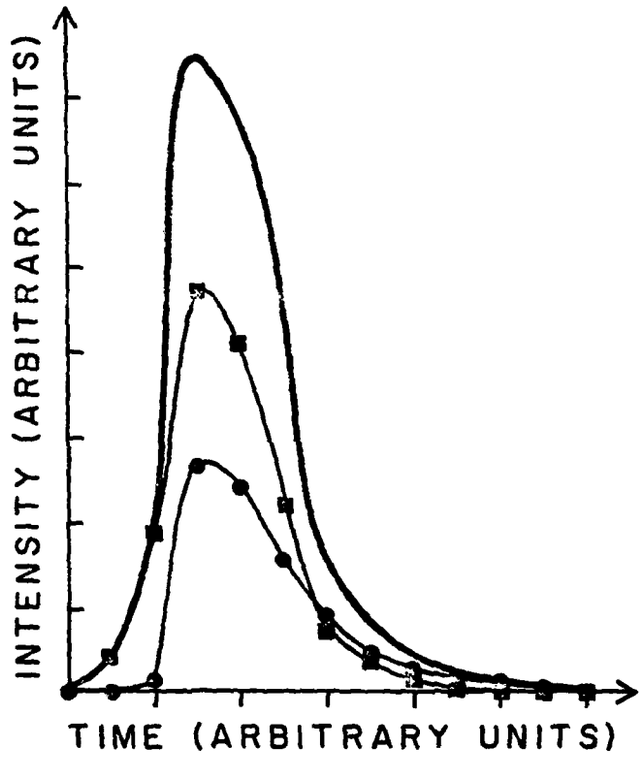
Fig. 4: The coelution of n-heptane and methylcyclohexane at:

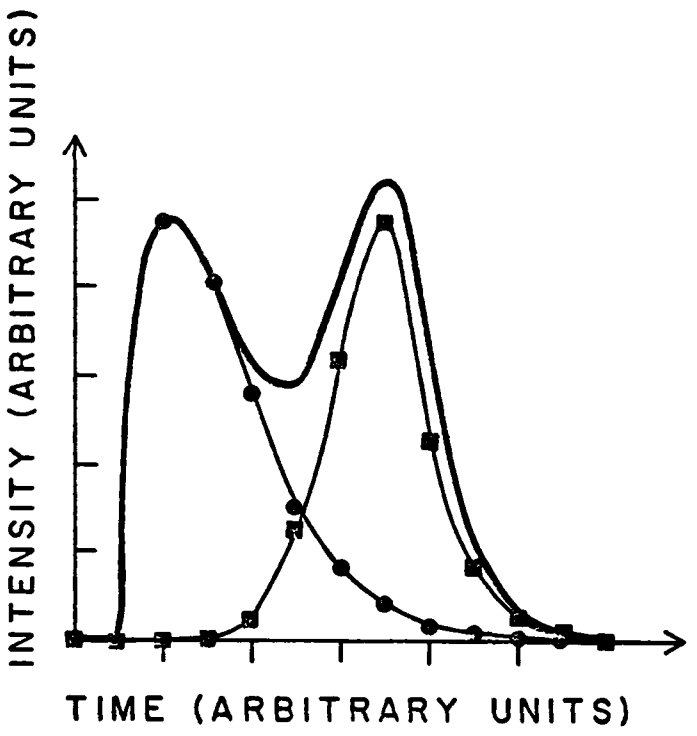
(a) 85°C and a flow rate of 33.2ml/minute,

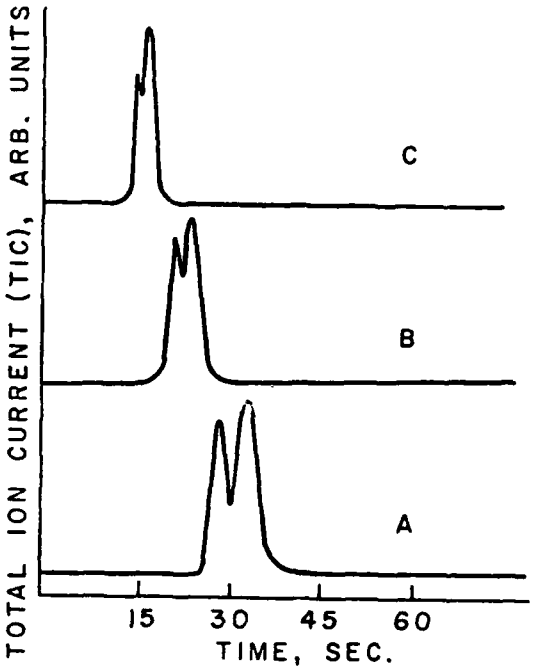
(b) 104°C and a flow rate of 33.3ml/minute and

(c) 130°C and a flow rate of 34ml/minute.


b
a
(a_{1u}, b_{1u})
(a_1, b_1)
(a_2, b_2)
(a_3, b_3)
(a_4, b_4)
(a_{2u}, b_{2u})


INTENSITY (ARBITRARY UNITS)
TIME (ARBITRARY UNITS)


INTENSITY (ARBITRARY UNITS)
TIME (ARBITRARY UNITS)


TOTAL ION CURRENT (TIC), ARB. UNITS
C
B
A
15
30
45
60
TIME, SEC.